International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S5, December 2019

Regression Heuristics by Optimal Tridimensional Features of Electrocardiogram for Arrhythmia Detection

S.Aarathi, S. Vasundra

Abstract: ***Computer aided predictive analytics are vital in noncommunicable diseases. In particular, early diagnosis of arrhythmia (heart related disease) is crucial to prevent sudden deaths due to heart failure. The critical context to prevent deaths caused by arrhythmia is early prediction of the arrhythmia scope. The clinical experts widely consider the Electro Cardio Gram (ECG) report as primary parameter to scale the scope of arrhythmia. However, the diagnosis accuracy of clinical experts is highly correlate on their expertise. Unlike the other domains, the sensitivity that is the accuracy in disease-prone is very much crucial in clinical practices. Particularly, the accuracy and sensitivity are more vital in computer-aided heart disease prediction methods. Hence, the recent research contributions are quantifying the possibilities of optimizing machine-learning approaches to achieve significance in computer-aided methods to perform predictive analysis on arrhythmia detection. Regarding this context, this manuscript is defining a Regression Heuristics by Tridimensional Features of the electrocardiogram reports, which has intended to perform arrhythmia prediction. The experimental study evincing the significance of the proposed model that scaled against the contemporary methods.***

Keywords: ***Predictive Analytics, Machine Learning, Electrocardiogram, Feature Optimization, Linear Regression, Classification.***

I. INTRODUCTION

Health epidemic conditions are among intrinsic issues encountered by nations across the world. Heart disease is one of the significant challenges that are prevalent for healthcare institutions. Early detection and timely medical support can help the patients recover from the issues of heart disease, and thousands of lives can be saved. and thousands of lives can be saved. ECG (Electro Cardio Gram) is the vividly used test for diagnosing the heart functional parameters at initial levels. The purpose of ECG is to understand the electric impulse condition of the heart and to identify any kind of abnormalities that are integral to the problem [1]. The signals from ECG are usually the outcome of T waves, QRS complex, and the P waves. The parameters that are essential for examining the case of heart patients are about shape, relationship amidst P wave, QRS complex, R-R interval, and T wave. Any kind of abnormalities over the parameters can lead to the ailment in terms of heart conditions, which might take place because of the distinct set of reasons [2].

Manuscript published on 30 December 2019.
***** Correspondence Author (s)
S.Aarathi, Reasearch Scholar, Department of Computer Science and Engineering, JNTUA, CEA Anantapuramu, Andrapradesh, India. Email: aarthis1@gmail.com
Dr.S. Vasundra, Professor, Department of Computer Science and Engineering, JNTUACEA Anantapuramu, Andrapradesh, India.

Irregularity in the case of the heart rhythms is about Arrhythmia and in certain cases, the results are an outcome of certain intrinsic heart ailment conditions, that could impact the survival of the patients. It is paramount importance that such problems are diagnosed at very early stages that can lead to better prevention opportunities. In medical terms, Arrhythmia is the scenario wherein the smooth functioning of the heart's electrical system is disturbed and it leads to impact on the heart in terms of disturbance in the pulse rates and skipping beats that could lead to nonsequential movement in the case of the heart signals. The process of identifying Arrhythmia is to understand the symptoms of blood pumping from heart, shortness in the breath, fatigue, unconsciousness, chest pain and the ECG conditions [3]. It can be classified into two broad categories like bradycardia and tachycardia. Bradycardia leads to slow beat pace in the heart wherein it is below the rate of 60 beats per minute (bpm), wherein the tachycardia leads to increment in the faster heartbeats that could rise to 100 bpm [4].

Contemporary range of healthcare solutions for cardiac treatments is leading to acknowledging the need for handling the arrhythmia classifications in more effective ways. In terms of developing the remote healthcare systems for cardiac patients, the importance is about handling the accurate diagnostic system, wherein the role of novel computer aided solutions like the machine learning solutions had proposed. The models that have proposed in the earlier studies indicate potential kind of solution for addressing the challenges of electrocardiogram signals [5] [6]. Choosing right kind of technique for classification of arrhythmia is complex as there are various parameters and contextual application of the model is essential in addition to the historic data analysis in general and more specific to the patient records.

In this paper, the impact discussed is about an efficient system that predicts the scope arrhythmia in the given electrocardiogram is positive or negative. The electrocardiograms labeled as positive (arrhythmia prone) and negative (normal) have used in learning phase constitutes features, which have fall in to one of the tridimensional categories listed as intervals, axis, and signals. Most of these features fail to signify the scope of the labels. In contrast, these insignificant features lead to increase in false alarming. Hence, it is obvious to select optimal features. The distribution diversity method called t-test [7] has used to determine the feature significance towards the labels positive, and negative.

These features further have used to build the regression heuristics, which have used further to perform the prediction analytics of the arrhythmia scope in electrocardiograms.

II. RELATED WORK

It is imperative from the literature that arrhythmia discrimination is the significant objective of the many contributions. The techniques such as linear discriminants (LD) [8], [9]. [10] or the decision tree-based as discussed in [10], [11], [12]. Few of the conventional set up of neural networks-based solutions [13-18], and in [14], [19], [20], [21], [22], [23] the SVM based training models were discussed. The conditional random field-based model is proposed in [24], and the deep learning models are used in [18], [25], [26], [27].

Many of the earlier works were more focused on identifying the best combination of features and sometimes in terms of developing complex signal processing models in terms of choosing the right kind of subset reduction conditions for handling the arrhythmia classification [28].

While in one hand, the choices are profoundly based on the input features that are considered based on the morphological features garnered from the time domain [8], [19], [20], [29], complex heartbeat representations [21], wavelet transforms [9], [13], [14], [15], [16], [23], [30], or higher order statistics (HOS) [9], [11], [12], [14] and frequency-domain features [11], [12], [21], [22], [31]. In the other dimension, the feature selection models like the independent component analysis (ICA) [23], [30], PSO (particle swarm optimization) [21] and the GA-BPNN (genetic algorithm back propagation neural networks) are used.

In the other research model executed, the process of automated artificial neural networks was proposed for classifying the arrhythmia patients, based on a standard 12-lead ECG data recording. Missing data is carried out by focusing on replacing attribute values along with the closet value based on the concern class. Post replacement of the missing values, a perceptron of multilayer constituting static backpropagation model is used for arrhythmia classification [32]. Like the works that have executed based on the generalized feed-forward kind of neural network solution [33], MLP constitutes one against different models [34] in terms of classifying the cardiac arrhythmia across different classes. A novel approach proposed in [35] relies on the correlation-based feature selection technique for choosing the appropriate kind of features using the UCI ECG datasets and over the incremental backpropagation neural networks alongside the Levenberg–Marquardt has employed to ensure early and more accurate detection of an arrhythmia.

Decision trees were used to handle the successful classification of the cardiac arrhythmia for designing a computer-assisted diagnosis system. Such diagnosis and decision support systems can be resourceful for physicians in handling the disease conditions more effectively and towards reducing the workload conditions in the healthcare institutions. Random forest ensemble method is used for resampling which is proposed for improving the classification of systems based on the arrhythmia detection condition [36]. In [37], numerous machine learning models were discussed which constitutes even the neural networks, gradient boosting, random forest and SVM which are used for classification based on the application of rigorous preprocessing and feature selection technique for the ECG data. Similarly, the SVM, KNN, Logistic regression, decision trees, OneR, Naïve Bayes and J48 are used in the studies for different classes used to classify the arrhythmia [38], [39], [40].

Contemporary work over the ECG datasets is carried out in [41] using the SVM based methods for detection of arrhythmia with a selection of features based on the principal component analysis [41], [42]. An effective model of classification for arrhythmia patients was proposed that relies on SVM and KNN for handling training model and to enhance accuracy measure, which is attained based on the combination of F-score and SFS (Sequential forward search) in handling the selection features [43].

The recent contributions “Fast Machine Learning Model (FMLM) [44] for ECG-Based Heartbeat Classification” and “Multiclass Classification of Cardiac Arrhythmia (MCCA) [45] Using Improved Feature Selection and SVM Invariants” have considered all possible features and variance in arrhythmia projection format (multiple classes) of the electrocardiograms in respective order. However, these contemporary methods still evincing the considerable false alarming and poor sensitivity, which is due to the improper feature adaption and ignorance of the feature dimensionality. In this contest, the contribution of this manuscript portrayed a tridimensional feature selection and optimization process, and built a scale by regression heuristics derived from optimal tridimensional features to perform predictive analysis of the electrocardiograms towards arrhythmia scope detection.

III. METHODS AND MATERIALS

This section explores the methods and materials related to data used, feature selection, feature optimization, learning and label prediction strategies proposed in this manuscript.

1.1 The data structure and features

An arrhythmia is a problem with the rate or rhythm of heartbeat, wherein the heartbeats have abnormal conditions of too high, too low or having the non-conducive pattern. The format of the data considering is the electrocardiography signal series, which is the output of the electrocardiogram that uses to identify the heart beat format. These signals further used to determine the diversified features, which have categorically grouped and referred further as tridimensional features. These dimensions and respective features have explored in following sections.

1.2 Intervals

RR Interval: The time elapsed among the sub sequential R-waves of QRS signal over the ECG and its reciprocal in terms of HR which is a function of intrinsic properties for the sinus node and the autonomic influence conditions.

PR Interval: It is the estimated time for onset of the P wave for starting the QRS complex. It reflects on the conduction based on the AV node. The normal PR interval ranges between 120-200ms (0.12-0.20s) in terms of duration. If the PR interval is >200ms, the presence of first-degree heart block is imperative.

QRS Duration: Usual duration of the QRS complex is around 0.08 to 0.10 seconds, which can be attributed as 80 and 100 milliseconds. If the duration is in the range of 0.10 to 0.12 seconds, it is considered as intermediate or the ones that are slightly prolonged. A QRS duration which is higher than that of 0.1 seconds is seen as abnormal conditions.

QT interval is seen as the intrinsic parameter as the multiple durations were reported. Usually, the general QT interval is around 400 to 440 milliseconds (ms) or the ones that are 0.4 to 0.44 seconds. In the case of female patients, it is observed that they have a longer QT interval than men. Lower heart rates also result in longer QT interval conditions.

QTc interval is about defining the normal QTc which varies in terms of being equal to or lower than that of 0.40 s, 0.41s (≤410ms), 0.42s (≤420ms) or 0.44s (≤440ms). In the case of sudden cardiac death, wherein the borderline QTc in the case of male patients are considered as 431-450ms, in the case of female patients it is seen as 470ms. "abnormal" QTc records in males is a QTc > 450 ms; and, in females, > 470 ms.

1.3 Axis

Axis of the ECG is among the key direction in terms of overall electrical activity for the heart. It can be normal or rightward or leftward or can even be in terms of indeterminate conditions like the northwest axis.

Degree of P wave Axis: It reflects on the atrial depolarization wherein the sinus node takes place, which is considered as a sinoatrial node, which creates an action which can depolarize the atria. The P wave must be upright in lead II when the action potential is generated from the SA node.

Degree of QRS Wave Axis: It is highly important to determine the QRS axis, wherein the normal QRS axis shall constitute in the range of -30 to +90 degrees. Left axis deviation can be termed as major QRS vector ranging -30 to -90 degrees. Right axis deviation takes place in the QRS axis in the range of +90 to +180 degrees.

Degree of T wave axis: In the ECG, the T wave indicates repolarization of ventricles. The interval right from the QRS complex to the apex of T wave is signified as an absolute refractory period. T wave is a kind of most labile wave in the ECG. T wave changes along with amplitude T waves and are abnormally inverted T waves that might be resulting outcome of a distinct set of cardiac and non-cardiac conditions. Normal T wave is usually in a similar direction as to the QRS except for the kind of right precordial leads.

But in the case of the P wave or T wave axis, usually, the measures indicate the QRS axis and the limb leads that are to be investigated. The usual QRS axis has to be in the range of -30 to +90 degrees. Left axis deviation can be stated as a major QRS vector that falls in the range of -30 to -90, whereas the right axis range is around +90 to +180. Hence, the indeterminate axis range have stated as $\{(+/-)190,-90\}$

1.4 Signal

Some of the statistical features of biomedical signals usually change based on time. Wavelet transform supports in signal representation wherein both time and frequency domains are considered, which will ensure there is capability in terms of analyzing the quasiperiodic signals like the ECG. Wavelet transform must be employed in terms of processing the ECG signals for feature extraction [46], heart-beat recognition [47], de-noising [48]. The proposed model, wherein the DWT is used for feature extraction technique has the scope of decomposing into low-frequency or high-frequency approximation components.

Among the vividly used wavelets, it provides orthogonality properties which are like Daubechies, Discrete Meyer, Symlets, Coiflets [49]. Every heartbeat is disintegrated based on the finite impulse response conditions for the Discrete Mayers wavelet transform frequency which ranges from fourth-level approximation sub-band which is around 011.25Hz and the frequency range for the level four detail sub-band is 11.2522 Hz. Coefficients for the two hundred count are garnered based on the wavelet features that are processing using ICA for handling dimensionality reduction. Six of the major ICA components were chosen across each of the two DWT sub-bands which results in 12 morphological features from two sub-bands.

1.5 Preprocessing

This phase discards the input electrocardiograms that are not engaged either of the labels positive (arrhythmia), negative (benign) in the given training corpus. Further extracts the features of formats listed as intervals CS, Axis AS, and signals FS of the each input electrocardiogram (electrocardiogram) of the given training corpus. The record of intervals $\{r \exists r \in CS\}$ represents the different intervals stacked in corresponding electrocardiogram. The record of axis values $\{ar \exists ar \in AS\}$ of the electrocardiogram represents the divergent axis formats. Similarly, derives the record of signal n-grams $\{fo \exists fo \in FS\}$ of each electrocardiogram of the given corpus.

1.6 Optimizing the n-grams

Sets arC_+, arC_- from the set AS, and the sets foC_+, foC_- from the set FS

Cross-Modelling-Feature-Optimization (nS, P_+, P_-) Begin // member function that performs n-gram feature optimization, which receives n-grams nS and respective positive (prone to arrhythmia) and negative (not evincing arrhythmia) records as sets P_+, P_-

---

Let an index 'i' being initialized to 1

For each index $\{i \exists i = 1, 2, ..., n\}$ Begin //having value in range of minimum pattern size 1 to max pattern size n

Extract i -gram interval n-grams from the set nS as a set nP_i

Foreach i -gram $\{np \exists np \in nP_i \wedge |np| \cong i\}$ Begin // Find the support of the n-gram pattern np in regard to the label positive, which is as follows

$$s_+^{np} = \left(\sum_{j=1}^{|P_+|} \{1 \exists np \subseteq r \wedge r \in P_+\} \right) * \left(|P_+| \right)^{-1}$$

// this denotes the ratio of source records having the n-gram pattern np as subset against the total number of records of intervals labeled as positive and listed in set P_+.

Move the support s_+^{np} to a vector v_+^i

Similarly, find the support of the interval n-gram csp in regard to the label negative, which is as follows

$$s_-^{np} = \left(\sum_{j=1}^{|P_-|} \{1 \exists np \subseteq r_j \wedge r_j \in P_-\} \right) * \left(|P_-| \right)^{-1}$$

// this denotes the ratio of records of Intervals having the interval n-gram np as subset against the total number of records of Intervals labeled as negative and listed in set P_-.

Then move the support s_-^{np} to a vector v_-^i

Further, scale the variance between the vectors v_+^i, v_-^i using competent distribution diversity assessment method called dual tailed t-test [7] (see sec 3.4).

If the t-score, and p-value [50] observed from the dual tailed t-test has used further to estimate the optimality of the $i-gram$ interval n-grams as follows.

The resultant p-value is greater than the given probability threshold $p\tau$, the $i-gram$ patterns are said to be optimal and move these $i-gram$ patterns to the set onP that represents the resultant optimal n-gram features.

The process explored in previously mentioned statements execute for each index 'i' value range from 1 to n.
Return the optimal n-gram features as a set onP

End

1.7 T-Test for Distribution Diversity Estimation

The distribution diversity has taken as parameter to estimate the optimality of the features listed as n-grams of intervals csP, n-gram Axis n-grams arP, and n-grams of the signal n-grams foP towards the labels positive and negative. The detailed exploration of the optimal feature selection has portrayed in aforesaid section (sec 1.2). The t-test is adapted to estimate the distribution diversity of the features towards positive and negative labels. This section details the method of performing t-test to identify the distribution diversity. From the contribution [7], the scheme that evaluates the distribution diversity, which has titled as t-test is in use for selecting the optimal features relevant to the both labels.

Here, t-test is included for selecting the optimal-features associated to both the positive & negative records of the corpus CS .

The diversity between any two distinct vectors can depict using T-score as,

$$v_{+}^{stdev} = \left(\sum_{i=1}^{|v_+|} \left(x_i - \langle v_+ \rangle \right)^2 \right) * \left(|v_+| - 1 \right)^{-1}$$

$$v_{-}^{stdev} = \left(\sum_{j=1}^{|v_-|} \left(x_j - \langle v_- \rangle \right)^2 \right) * \left(|v_-| - 1 \right)^{-1} \quad \text{...(Eq 1)}$$

$$t\_stdev = \sqrt{v_{+}^{stdev} + v_{-}^{stdev}}$$

$$t - score = \left(\langle v_+ \rangle - \langle v_- \rangle \right) * \left(t\_stdev \right)^{-1}$$

- In (Eq 1), the notations $\langle v_+ \rangle, \langle v_- \rangle$ entails the mean of the respective vectors v_+, v_-
- The notations x_i, x_j refer the entries of the vectors in sequence from the index 1 to vector sizes $|v_+|, |v_-|$ represented by i, j in respective order.

Ratio between the absolute difference of the means $\langle v_+ \rangle - \langle v_- \rangle$ of corresponding vectors v_+, v_- and the aggregate of the deviations observed from the corresponding vectors.

Then compute p-value [50] (degree-of probability) in the t-table [51] for attained t-score. Here, the p-value, which is lower than probability threshold $p\tau$ signifies both the vectors are different; therefore, the patterns represented by the entries of the corresponding vectors has said to be optimal-feature.

1.8 Optimizing n-gram Intervals

For each record of intervals $\{r \exists r \in CS\}$ of the set CS , find all possible interval n-grams of size 1 to record-size $|r|$. Move all possible interval n-grams of count $(|r|+1)*(|r|)*2^{-1}$ discovered from the record $\{r \exists r \in CS\}$ to the set csP . The set CS has to partition in to two sets such that one represents all the records of the set having positive label, and the other set represents the records of negative label. This phase discovers the sets CS_+, CS_- form the set CS . Further, invokes the member function Optimizing N-grams (csP, CS_+, CS_-) that returns optimal interval n-grams, which has been received as $ocsP$

1.9 Optimizing the Axis n-grams

Prepare a corpus arC that contains set of records, such that each record $\{ar \exists ar \in arC\}$ contains the values of the Axis n-grams exists in electrocardiogram record $\{r \exists r \in CS\}$ of the corpus CS and entails the label assigned to the corresponding electrocardiogram record $\{r \exists r \in CS\}$. Further, partition the corpus arC in to two sets $arC_+, and\ arC_-$, such that the set arC_+ contains the records exists in corpus arC and having the label positive, and the other set arC_- contains the records of the corpus arC that are having label negative. Each record $\{ar \exists ar \in arC\}$ of the set arC with index i represents the Axis n-grams of the record $\{r \exists r \in CS\}$ with index i in the corpus CS . The similar process that stated in above section (sec 3.4) has used to list the total axis n-grams from each electrocardiogram as a set arP .Further, invoke the function called "Cross-Modelling-Feature-Optimization (arP, arC_+, arC_-)" (see sec 3.3) that returns optimal interval n-grams, which have been received as set $oarP$.

1.10 Optimizing the signal n-grams

The process that follows to list the signal n-grams is very much similar to the process of listing axis n-grams. Further, list the signal n-grams as a set foP . Prepare the corpus foC that contains set of records, such that each record $\{fr \exists fr \in foC\}$ contains the signal n-grams, resulting from the electrocardiogram as record $\{r \exists r \in CS\}$ of the corpus CS and entails the label assigned to the corresponding record $\{r \exists r \in CS\}$. Each record $\{fr \exists fr \in foC\}$ of the set foC with index $'i'$ represents the values of the signal n-grams exhibited by the electrocardiogram record $\{r \exists r \in CS\}$ with index $'i'$ in the corpus CS .

Further, partition the corpus foC in to two sets $foC_+, and\ foC_-$, such that the set foC_+ contains the records exists in corpus foC and having the label positive, and the other set foC_- contains the records of the corpus foC that are having label negative. Further, the similar version of optimal n-gram selection has adapted that invokes "Optimizing-Ngrams (foP, foC_+, foC_-)" (see sec 3.3) that returns optimal n-grams, which has to receive as a set $ofoP$.

1.11 Regression Heuristics by N-gram Coefficients

This section delivers the N-gram Coefficients for diversified features listed optimal n-grams, which have discovered from their empirical probabilities of the n-grams.

Find-Feature-Coefficients (nP) **begin**	// the function that discovers N-gram Coefficients of the feature formats
$ss = \left(\sum_{j=1}^{\lvert nP \rvert} \{ s(np_j) \exists np_j \in nP \} \right) * \lvert nP \rvert^{-1}$	// Find the mean of the empirical probabilities of the optimal n-grams as n-grams coefficient
$ssd = \left(\sum_{j=1}^{\lvert nP \rvert} \left\{ \sqrt{(ss - s(np_j))^2} \exists np_j \in nP \right\} \right) * \lvert nP \rvert^{-1}$	//Finding the root mean square distance (mean deviation) of the empirical probability of the optimal n-grams.
$ssl = ss - ssd$	// the absolute distance of the empirical probability and respective mean deviation is the lower bound of the n-grams coefficient
$ssu = ss + ssd$	// the aggregate of the empirical probability and respective mean deviation is the upper bound of the n-grams coefficient
Return (ss, ssl, ssu)	// returns the n-grams coefficient, respective lower and upper bounds as Regression Heuristics
End	

1.11.1 Regression Heuristics of the Interval n-grams

The N-gram Coefficients of the intervals csP_+ have derived by invoking the function **Find-Feature-Coefficients** (csP_+), which returns the N-gram Coefficients $(ss_+^{csP}, ssl_+^{csP}, ssu_+^{csP})$ as Regression Heuristics of the interval n-grams of the positive label. Similarly, the Regression Heuristics $(ss_-^{csP}, ssl_-^{csP}, ssu_-^{csP})$ of the interval n-grams of negative label have to estimate by invoking the function Find-Feature-Coefficients (csP_-), which passes the interval n-grams csP_- of the negative label as input parameter

1.11.2 Regression Heuristics of the Axis n-grams

The set arP_+ Axis n-grams of the positive label have to pass as input parameter of the function Find-Feature-Coefficients (arP_+) that return a set of coefficients $(ss_+^{arP}, ssl_+^{arP}, ssu_+^{arP})$ as Regression Heuristics. Similarly, the Regression Heuristics $(ss_-^{arP}, ssl_-^{arP}, ssu_-^{arP})$ of the Axis n-grams of negative label have to estimate by invoking the function Find-Feature-Coefficients (arP_-) with axis n-grams arP_- of the negative label as input parameter

1.11.3 Regression Heuristics of the signal n-grams

The N-gram Coefficients $(ss_+^{foP}, ssl_+^{foP}, ssu_+^{foP})$ of the signal n-grams of the positive label as Regression Heuristics have to receive from the function Find-Feature-Coefficients (foP_+), which has to invoke by passing the set of signal n-grams foP_+ of the positive label. The Regression Heuristics $(ss_-^{foP}, ssl_-^{foP}, ssu_-^{foP})$ of the signal n-grams of the negative label has emerged as the return set of the function Find-Feature-Coefficients (foP_-), which has to invoke by passing the signal n-grams of the set foP_- as input parameter.

1.11.4 Label Prediction

For given unlabeled electrocardiogram r, derive all possible interval n-grams, which buffered as a set $csPT$. Further, extract the Axis n-grams of the electrocardiogram r and list them as a set $arPT$. Similarly, determine all possible signal n-grams of the given electrocardiogram as a list foT.

Discover the interval n-grams those are common in both the sets $csP, and\ csPT$ as a set $ccsp$. Then find the positive and negative confidence of the interval n-grams listed in the set $ccsp$ as follows in (Eq 2, 3, 4).

$$ccsp = csP \bigcap csPT \quad ...(Eq\ 2)$$

// discover the interval n-grams those are common in both the sets $csP, and\ csPT$ as a set $ccsp$

$$\text{csc}_+ = \left(\sum_{i=1}^{\lvert ccsp \rvert} \{ s_+^{csp_i} \exists csp_i \in ccsp \} \right) * m^{-1} \quad ...(Eq\ 3)$$

// Find the confidence of the interval n-grams for positive label, which is the ratio of empirical probabilities of the respective records of Intervals labeled as positive of the training corpus

$$\text{csc}_- = \left(\sum_{i=1}^{\lvert ccsp \rvert} \{ s_-^{csp_i} \exists csp_i \in ccsp \} \right) * m^{-1} \quad ...(Eq\ 4)$$

// Find the confidence of the interval n-grams of negative label, which is the ratio of empirical probabilities of the respective electrocardiogram labeled as negative of the training corpus.Similarly, find the positive and negative confidence of the Axis n-grams, and signal n-grams listed in the sets $arPT$, $foPT$ as follows in (Eq 5, 6, 7).

$$carp = arPT \bigcap arP \quad ...(Eq\ 5)$$

// Find the patterns common in both sets $arPT, arP$

$$arc_{+} = \left(\sum_{i=1}^{|carp|} \left\{ s_{+}^{arp_i} \exists arp_i \in carp \right\} \right) * p^{-1} \quad ...(Eq\ 6)$$

// Find the confidence of the axis n-grams for positive label, which is the ratio of empirical probabilities of the respective Axis n-grams obtained from the electrocardiograms labeled as positive of the training corpus.

$$arc_{-} = \left(\sum_{i=1}^{|carp|} \left\{ s_{-}^{arp_i} \exists arp_i \in carp \right\} \right) * p^{-1} \quad ...(Eq\ 7)$$

// Find the confidence of the Axis n-grams for the negative label, which is the ratio of empirical probabilities of the respective Axis n-grams obtained from the electrocardiograms labeled as negative of the training corpus. Further phase finds the positive and negative confidence of the signal n-grams as follows in (Eq 8, 9, 10).

$$cfop = foPT \cap foP \quad ...(Eq\ 8)$$

// Find the patterns common in both sets $foPT, foP$

$$foc_{+} = \left(\sum_{i=1}^{|cfop|} \left\{ s_{+}^{fop_i} \exists fop_i \in cfop \right\} \right) * q^{-1} \quad ...(Eq\ 9)$$

// Find the confidence of signal-Ngrams for positive label, which is the ratio of empirical probabilities of the respective signal n-grams obtained from the electrocardiograms labeled as positive of the training corpus.

$$foc_{-} = \left(\sum_{i=1}^{|cfop|} \left\{ s_{-}^{fop_i} \exists fop_i \in cfop \right\} \right) * q^{-1} \quad ...(Eq\ 10)$$

// Find the confidence of signal n-grams for negative label, which is the ratio of empirical probabilities of the respective signal n-grams obtained from the electrocardiograms labeled as negative of the training corpus. Further, these confidence metrics has to correlate to identify the label, which is as follows

- Label the given test electrocardiogram as positive in regard to regression heuristics of interval n-grams, if the
  - $\text{csc}_{+} \geq ssu_{+}^{csP}$ //positive confidence csc_{+} of the Interval n-grams of the given test electrocardiogram is greater than the upper bound of the regression heuristics ssu_{+}^{csP} (see sec 3.8.1) of the interval n-grams csP for positive label.
  - $\text{csc}_{+} \geq ss_{+}^{csP} \,\&\&\, \text{csc}_{-} < ss_{-}^{csP}$ // positive confidence csc_{+} of the interval n-grams of the given test electrocardiogram is greater than the regression heuristics ss_{+}^{csP} of the interval n-grams for positive label, and negative confidence csc_{-} of the interval n-grams of the test electrocardiogram is less than the regression heuristics ss_{-}^{csP} of the interval n-grams for negative label.
  - $\text{csc}_{+} \geq ssl_{+}^{csP} \,\&\&\, \text{csc}_{-} < ssl_{-}^{csP}$ //positive confidence csc_{+} of the interval n-grams of the given test electrocardiogram is greater than or equal to the regression heuristics lower-bound ssl_{+}^{csP} of the interval n-grams for positive label, and negative confidence csc_{-} of the interval n-grams of the given test electrocardiogram is less than the lower-bound of the regression heuristics ssl_{-}^{csP} of the interval n-grams for negative label.
- Label the given test electrocardiogram as positive in regard to axis regression heuristics, if the
  - $arc_{+} \geq ssu_{+}^{arP}$ //positive confidence arc_{+} of the Axis n-grams of the given test electrocardiogram is greater than the upper bound of the regression heuristics ssu_{+}^{arP} of the Axis n-grams for positive label (see sec 3.8.2).
  - $arc_{+} \geq ss_{+}^{arP} \,\&\&\, arc_{-} < ss_{-}^{arP}$ //positive confidence arc_{+} of the Axis n-grams of the given test electrocardiogram is greater than the regression heuristics ss_{+}^{arP} of the Axis n-grams for positive label, and negative confidence arc_{-} of the Axis n-grams of the given test electrocardiogram is less than the regression heuristics of the Axis n-grams for negative label.
  - //positive confidence of the Axis n-grams of the given test electrocardiogram is greater than or equal to the regression heuristics lower-bound of the Axis n-grams for positive label, and negative confidence of the Axis n-grams of the given test electrocardiogram is less than the lower-bound of the regression heuristics of the Axis n-grams for negative label.
- Label the given test electrocardiogram as positive in regard to signal regression heuristics, if the
  - //positive confidence of the signal n-grams of the given test electrocardiogram is greater than the upper bound of the regression heuristics of the signal n-grams for positive label (see sec 3.8.3).
  - //positive confidence of the signal n-grams of the given test electrocardiogram is greater than the regression heuristics of the signal n-grams for positive label, and negative confidence of the signal n-grams of the given test electrocardiogram is less than the regression heuristics of the signal n-grams for negative label.
  - //positive confidence of the signal n-grams of the given test electrocardiogram is greater than or equal to the regression heuristics lower-bound of the signal n-grams for positive label, and negative confidence of the signal n-grams of the given test electrocardiogram is less than the lower-bound of the regression heuristics of the signal n-grams for negative label.
- Label the given test electrocardiogram as negative in regard to regression heuristics of interval n-grams, if the,
  - positive confidence of the interval n-grams of the given test electrocardiogram is less than the regression heuristics of the interval n-grams for positive label and the negative confidence (of the interval n-grams of the test electrocardiogram) is greater than the upper bound of the regression heuristics of the interval n-grams (see sec 3.8.2).

- positive confidence of the interval n-grams of the given test electrocardiogram is less than the lower-bound of the regression heuristics of the interval n-grams for positive label, and the negative confidence (of the interval n-grams of the test electrocardiogram) is greater than the regression heuristics (of the interval n-grams) for negative label (see sec 3.8.2).

- Label the given test electrocardiogram as negative in regard to axis regression heuristics, if the,
  - //positive confidence of the Axis n-grams of the given test electrocardiogram is less than or equals the regression heuristics of the Axis n-grams for positive label and the negative confidence(of the Axis n-grams of the test electrocardiogram) is greater than the upper bound of the regression heuristics of the Axis n-grams for negative label (see sec 3.8.2).
  - //positive confidence of the Axis n-grams (of the given test electrocardiogram) is less than the lower-bound of the regression heuristics of the Axis n-grams for positive label and the negative confidence(of the Axis n-grams of the test electrocardiogram) is greater than the regression heuristics (of the Axis n-grams) for negative label (see sec 3.8.2).
- Label the given test electrocardiogram as negative in regard to signal regression heuristics, if the,
  - //positive confidence of the signal n-grams of the given test electrocardiogram is less than or equals the regression heuristics of the signal n-grams for positive label and the negative confidence(of the signal n-grams of the test electrocardiogram) is greater than the upper bound of the regression heuristics of the signal n-grams for negative label (see sec 3.8.3).
  - //positive confidence of the signal n-grams (of the given test electrocardiogram) is less than the lower-bound of the regression heuristics of the signal n-grams for positive label and the negative confidence (of the signal n-grams of the test electrocardiogram) is greater than the regression heuristics (of the signal n-grams) for negative label (see sec 3.8.3).

In contrast to the above conditions, the given record of Intervals has treated as suspicious. However, the definition of these conditions is solely domain sensitivity specific.

IV. EMPIRICAL STUDY

This section explores the empirical study that carried on proposed model and the other contemporary models using benchmark dataset. The performance significance of the proposed model has scaled by comparing the observed results of the classification assessment metrics and they are "Precision, Specificity, Sensitivity, Accuracy, F-measure, False Alarming, and Matthews's correlation coefficient (MCC)". The results evinced from the proposed model RH-TOF, and the other contemporary models "A fast machine learning model (FMLM)" [44] and "Multiclass Classification of Cardiac Arrhythmia (MCCA) to Predict Advanced Arrhythmia" [45] has compared and concluded the significance of the proposed model towards arrhythmia detection.

1.11.5 The dataset

MIT-BIH database [52] is the first standard test material that is available for evaluating arrhythmia detection. The database chosen has widely used in the cardiac dynamics related research worldwide. The dataset constitutes 48 half-hour excerpts of two-channel and 24-hour ECG recordings garnered from 47 subjects which are studied by BIH Arrhythmia Labs. The 48 half-hour excerpts were split into two groups.

- 23 (the "100 series") are chosen in random using a collection of over 4000 Holter tapes, and the other
- 25 (the "200 series") are chosen for including examples of rare but clinically crucial arrhythmias that would not be well exemplified in a small random sample.

V. RESULTS DISCUSSION

ECG signals garnered from Physionet MIT-BIH arrhythmia database is used for analysis process, wherein the text header file is shown as (.hea) and binary file as (.dat) and the (.atr) as binary annotation file (.hea) is the short text file that reflects on the contents of the signals that includes a number of samples, signal format, detailed clinical information, type of signal, etc. All the records constitute (.dat) files which comprise digitized samples for one or more signals stored in 212 formats and more records constitute one or more levels of (.atr). Annotation files constitute the labels set, wherein every set detail the feature for one or more signals during the recorded time period.

Table 1: Average and standard deviation for the various metrics

	Precision	Specificity	Sensitivity	Accuracy	F-measure	False alarming	MCC
RH-TOF	0.888±0.006	0.866±0.007	0.831±0.009	0.846±0.008	0.877±0.006	0.154±0.008	0.692± 0.015
MCCA	0.858±0.009	0.829±0.011	0.805±0.006	0.815±0.008	0.843±0.01	0.185±0.008	0.63±0 .017
FMLM	0.866±0.006	0.843±0.008	0.79±0.005	0.813±0.006	0.855±0.007	0.187±0.006	0.629± 0.012

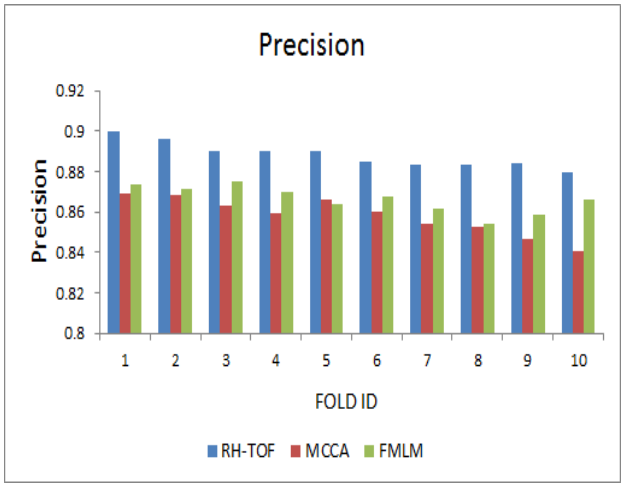

Figure 1: Graphical representation of Precision at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

The metric precision indicates the ratio of records labeled correctly as positive to the total amount of falsely labeled records as positive. Figure 1 represents the graph between 10-folds and precision for MCCA, FMLM, and RH-TOF. Average precision for the proposed method RH-TOF that perceived from the 10-fold strategy is 0.888±0.006. While the average precision for the contemporary methods MCCA & FMLM are 0.858±0.009& 0.866±0.006 in respective order. From the statistics as shown in Table 1, it is noticed that the proposed model RH-TOF performs better when compared to contemporary MCCA & FMLM methods.

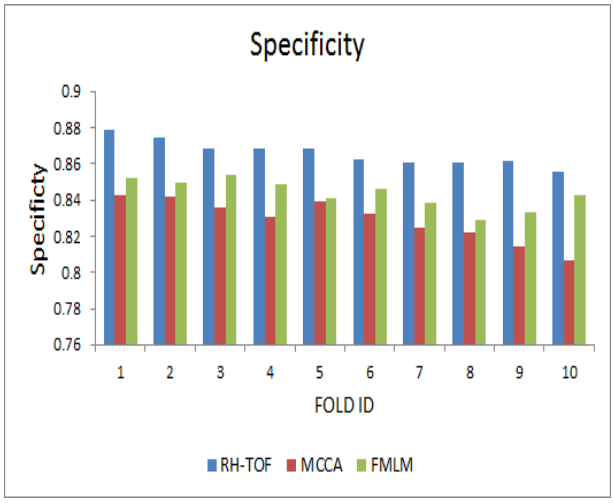

Figure 2: Graphical representation of specificity at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

The metric specificity denotes the ratio of correctly labeled records as negative against to total amount of negative label records. Figure 2 represents the graph between 10-folds and specificity for MCCA, FMLM, and RH-TOF. Average specificity for the proposed method RH-TOF and contemporary methods MCCA & FMLM that observed from 10-fold scheme are 0.866±0.007, 0.829±0.011, 0.843±0.008 respectively. From the statistics, it is noticed that the proposed model RH-TOF performs better when compared to contemporary MCCA & FMLM methods.

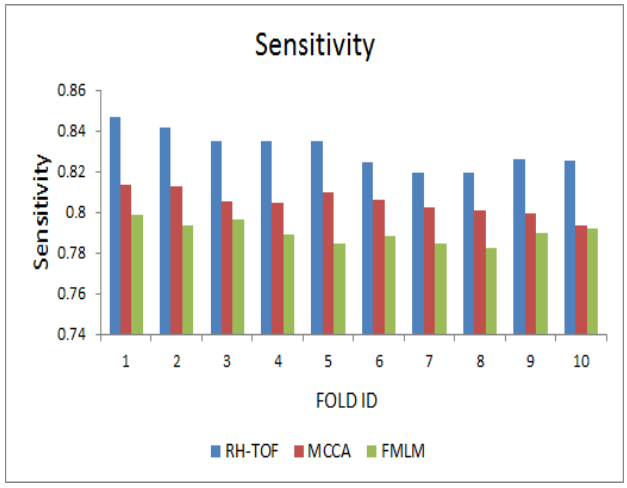

Figure 3: Graphical representation of sensitivity at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

The metric sensitivity denotes ratio of the test electrocardiograms labeled correctly as positive in averse to total amount of the positive records provided for testing. Figure 3 depicts the graph between 10-folds and sensitivity for MCCA, FMLM, and RH-TOF Average sensitivity for the proposed method RH-TOF that perceived from 10-fold scheme is 0.831±0.009. While the average sensitivity for the contemporary methods MCCA & FMLM are 0.805±0.006 & 0.79±0.005 in respective order. From the statistics, it is noticed that the proposed model RH-TOF is more significant when compared to contemporary MCCA & FMLM methods.

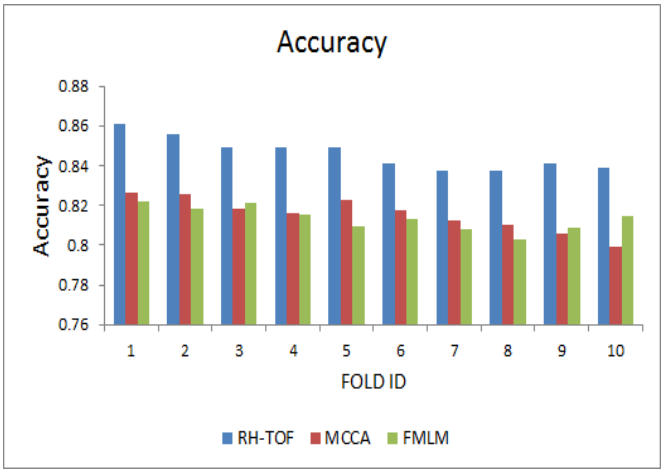

Figure 4: Graphical representation of Accuracy at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

The metric accuracy indicates the total performance for selecting the labels of specified unlabeled records that is ratio of cumulative of correctly labeled positive & negative records against to the overall amount of records by both the labels provided for testing. Figure 4 signifies the graph between 10-folds and accuracy for MCCA, FMLM, and RH-TOF. Average Accuracy for the

proposed method RH-TOF and contemporary methods MCCA & FMLM that perceived from 10-fold scheme are 0.846±0.008, 0.815±0.008, 0.813±0.006 respectively. From the statistics, it is noticed that the proposed model RH-TOF is considerably optimal when compared to contemporary MCCA & FMLM methods.

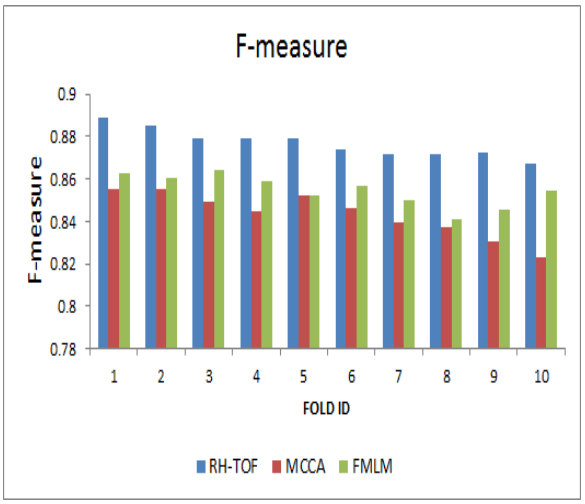

Figure 5: Graphical representation of F-measure at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

The metric F-measure indicates the weighted harmonic mean of precision & recall. Figure 5 depicts the graph between 10-folds and F-Measure for MCCA, FMLM, and RH-TOF. Average F-Measure for the proposed method RH-TOF that perceived from 10-fold scheme is 0.877±0.006. While the average F-measure for the contemporary methods MCCA & FMLM are 0.843±0.01 & 0.855±0.007 in respective order. From the statistics, it is noticed that the proposed model RH-TOF is more significant when compared to contemporary FMLM & MCCA methods.

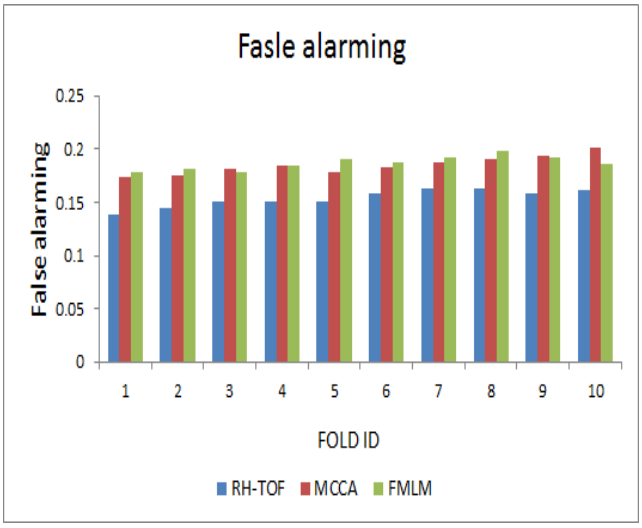

Figure 6: Graphical representation of false alarming rate at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods

Figure 6 signifies the graph between 10-folds and false alarming rate for MCCA, FMLM, and RH-TOF. The average false alarming rate for the proposed method RH-TOF and contemporary methods MCCA & FMLM that perceived from 10-fold scheme are 0.154±0.008, 0.185±0.008, 0.187±0.006 respectively. From the statistics, it is noticed that the proposed model RH-TOF is optimal with minimal false alarming that compared to contemporary MCCA & FMLM methods.

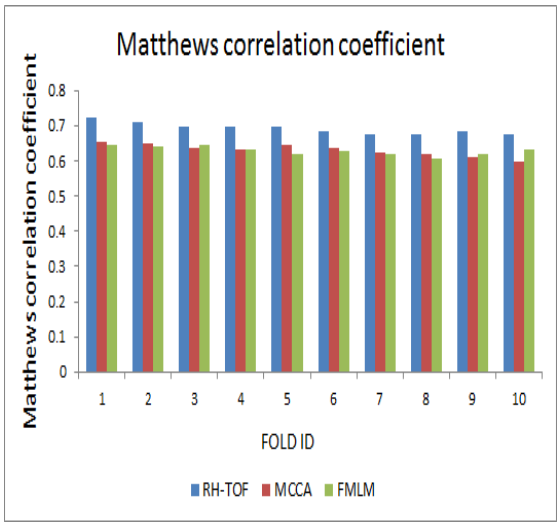

Figure 7: Graphical representation of Matthews's correlation coefficient at 10-folds for the proposed RH-TOF method and contemporary MCCA & FMLM methods.

The metric Matthews Correlation Coefficient (MCC) is utilized in machine learning as a measure of quality of binary classifications. Figure 7 depicts the graph between 10-folds and MCC for MCCA, FMLM, and RH-TOF. Average MCC for the proposed method RH-TOF that perceived from 10-fold scheme is 0.692±0.015. While the average MCC for the contemporary methods MCCA & FMLM are 0.63±0.017 & 0.629±0.012 respectively. From the statistics, it is noticed that the proposed model RH-TOF is more significant towards binary classification process that compared to contemporary MCCA & FMLM methods.

VI. CONCLUSION

Machine learning based predictive analysis for arrhythmia detection from the electrocardiogram has portrayed in this manuscript. In this regard, the tridimensional features listed as intervals, axis properties, and signal properties of the electrocardiogram (ECG) have used to train the proposed model, which is resulting regression heuristics of the each dimension of the tridimensional features adapted. These regression heuristics have further used to predict the given electrocardiogram is prone to arrhythmia (positive) or not (negative). The experimental study carried on the benchmark dataset MIT-BIH [52]. The performance significance of the proposal has scaled through outcomes of the statistical metrics such as precision, sensitivity, specificity, and accuracy, which have compared with the results obtained from the contemporary methods of the recent literature. The future research can drive confidently to derive soft computing methods, which can use these regression heuristics as fitness function.

On other dimension, the predictive analytics would define by using the other features such as gene expressions, electronic health and medical records.

REFERENCES

1. Zuo WM, Lu WG, Wang KQ, Zhang H. Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier. In2008 Computers in Cardiology 2008 Sep 14 (pp. 253-256). IEEE.
2. Alickovic E, Subasi A. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. Journal of medical systems. 2016 Apr 1;40(4):108.
3. Kumar SU, Inbarani HH. Neighborhood rough set-based electrocardiogram signal classification for diagnosis of cardiac diseases. Soft Computing. 2017 Aug 1;21(16):4721-33.
4. Özbay Y, Karlik B. A recognition of electrocardiogram arrhythmias using artificial neural networks. Seluk univ konya (turkey) electrical and electronics engineering; 2001 Oct 25.
5. Al Rahhal MM, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. Information Sciences. 2016 Jun 1;345:340-54.
6. NAYAK CG, Seshikala G, Desai U, Nayak SG. Identification of arrhythmia classes using machine-learning techniques. International Journal of Biology and Biomedicine. 2016;1:48-53.
7. Budak H, Taşabat SE. A modified t-score for feature selection. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik. 2016 Dec 20;17(5):845-52.
8. De Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using electrocardiogram morphology and heartbeat interval features. IEEE transactions on biomedical engineering. 2004 Jun 21;51(7):1196-206.
9. Mar T, Zaunseder S, Martínez JP, Llamedo M, Poll R. Optimization of electrocardiogram classification by means of feature selection. IEEE transactions on Biomedical Engineering. 2011 Feb 10;58(8):2168-77.
10. Krasteva V, Jekova I, Leber R, Schmid R, Abächerli R. Superiority of classification tree versus cluster, fuzzy and discriminant models in a heartbeat classification system. PloS one. 2015 Oct 13;10(10): e0140123.
11. Qurraie SS, Afkhami RG. electrocardiogram arrhythmia classification using time frequency distribution techniques. Biomedical engineering letters. 2017 Nov 1;7(4):325-32.
12. Afkhami RG, Azarnia G, Tinati MA. Cardiac arrhythmia classification using statistical and mixture modeling features of electrocardiogram signals. Pattern Recognition Letters. 2016 Jan 15; 70:45-51.
13. Dokur Z, Ölmez T. electrocardiogram beat classification by a novel hybrid neural network. Computer methods and programs in biomedicine. 2001 Sep 1;66(2-3):167-81.
14. Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T. Arrhythmia recognition and classification using combined linear and nonlinear features of electrocardiogram signals. Computer methods and programs in biomedicine. 2016 Apr 1; 127:52-63.
15. Martis RJ, Acharya UR, Min LC. electrocardiogram beat classification using PCA, LDA, ICA and discrete wavelet transform. Biomedical Signal Processing and Control. 2013 Sep 1;8(5):437-48.
16. Inan OT, Giovangrandi L, Kovacs GT. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE transactions on Biomedical Engineering. 2006 Nov 20;53(12):2507-15.
17. Javadi M, Ebrahimpour R, Sajedin A, Faridi S, Zakernejad S. Improving electrocardiogram classification accuracy using an ensemble of neural network modules. PLoS one. 2011 Oct 26;6(10):e24386.
18. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific electrocardiogram classification by 1-D convolutional neural networks. IEEE Transactions on Biomedical Engineering. 2015 Aug 14;63(3):664-75.
19. Zhang Z, Dong J, Luo X, Choi KS, Wu X. Heartbeat classification using disease-specific feature selection. Computers in biology and medicine. 2014 Mar 1;46:79-89.
20. Huang H, Liu J, Zhu Q, Wang R, Hu G. A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals. Biomedical engineering online. 2014 Dec;13(1):90.
21. Garcia G, Moreira G, Menotti D, Luz E. Inter-patient ecg heartbeat classification with temporal vcg optimized by pso. Scientific reports. 2017 Sep 5;7(1):10543.
22. Raj S, Ray KC, Shankar O. Cardiac arrhythmia beat classification using DOST and PSO tuned SVM. Computer methods and programs in biomedicine. 2016 Nov 1;136:163-77.
23. Ye C, Kumar BV, Coimbra MT. Heartbeat classification using morphological and dynamic features of electrocardiogram signals. IEEE Transactions on Biomedical Engineering. 2012 Aug 15;59(10):2930-41.
24. De Lannoy G, François D, Delbeke J, Verleysen M. Weighted conditional random fields for supervised interpatient heartbeat classification. IEEE Transactions on Biomedical Engineering. 2011 Oct 10;59(1):241-7.
25. Wu Z, Ding X, Zhang G. A novel method for classification of electrocardiogram arrhythmias using deep belief networks. International Journal of Computational Intelligence and Applications. 2016 Dec 25;15(04):1650021.
26. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, San Tan R. A deep convolutional neural network model to classify heartbeats. Computers in biology and medicine. 2017 Oct 1; 89:389-96.
27. Huanhuan M, Yue Z. Classification of electrocardiogram signals with deep belief networks. In2014 IEEE 17th International Conference on Computational Science and Engineering 2014 Dec 19 (pp. 7-12). IEEE.
28. Li H, Yuan D, Ma X, Cui D, Cao L. Genetic algorithm for the optimization of features and neural networks in electrocardiogram signals classification. Scientific reports. 2017 Jan 31;7: 41011.
29. Ortín S, Soriano MC, Alfaras M, Mirasso CR. Automated real-time method for ventricular heartbeat classification. Computer methods and programs in biomedicine. 2019 Feb 1;169: 1-8.
30. Ye C, Kumar BV, Coimbra MT. An automatic subject-adaptable heartbeat classifier based on multiview learning. IEEE journal of biomedical and health informatics. 2015 Aug 13;20(6):1485-92.
31. Zidelmal Z, Amirou A, Ould-Abdeslam D, Merckle J. electrocardiogram beat classification using a cost sensitive classifier. Computer methods and programs in biomedicine. 2013 Sep 1;111(3):570-7.
32. Jadhav SM, Nalbalwar SL, Ghatol AA. Arrhythmia disease classification using artificial neural network model. In2010 IEEE International Conference on Computational Intelligence and Computing Research 2010 Dec 28 (pp. 1-4). IEEE.
33. Jadhav SM, Nalbalwar SL, Ghatol AA. Generalized feedforward neural network based cardiac arrhythmia classification from electrocardiogram signal data. In2010 6th International Conference on Advanced Information Management and Service (IMS) 2010 Dec (pp. 351-356). IEEE.
34. Raut RD, Dudul SV. Arrhythmias classification with MLP neural network and statistical analysis. In2008 First International Conference on Emerging Trends in Engineering and Technology 2008 Jul 16 (pp. 553-558). IEEE.
35. Mitra M, Samanta RK. Cardiac arrhythmia classification using neural networks with selected features. Procedia Technology. 2013 Jan 1;10:76-84.
36. Özçift A. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. Computers in biology and medicine. 2011 May 1;41(5):265-71.
37. Batra A, Jawa V. Classification of arrhythmia using conjunction of machine learning algorithms and electrocardiogram diagnostic criteria. Training Journal. 1975 Sep.
38. Soman T, Bobbie PO. Classification of arrhythmia using machine learning techniques. WSEAS Transactions on computers. 2005 Jun;4(6):548-52.
39. Fazel A, Algharbi F, Haider B. Classification of Cardiac Arrhythmias Patients. CS229 Final Project Report. 2014.
40. Samad S, Khan SA, Haq A, Riaz A. Classification of arrhythmia. International Journal of Electrical Energy. 2014 Mar;2(1):57-61.
41. Kohli N, Verma NK. Arrhythmia classification using SVM with selected features. International Journal of Engineering, Science and Technology. 2011;3(8):122-31.
42. Kohli N, Verma NK, Roy A. SVM based methods for arrhythmia classification in electrocardiogram. In2010 international conference on computer and communication technology (ICCCT) 2010 Sep 17 (pp. 486-490). IEEE.
43. Niazi KA, Khan SA, Shaukat A, Akhtar M. Identifying best feature subset for cardiac arrhythmia classification. In2015 Science and Information Conference (SAI) 2015 Jul 28 (pp. 494-499). IEEE.

44. Alfaras M, Soriano MC, Ortín S. A fast machine learning model for electrocardiogram-based heartbeat classification and arrhythmia detection. Frontiers in Physics. 2019; 7:103.
45. Mustaqeem A, Anwar SM, Majid M. Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants. Computational and mathematical methods in medicine. 2018;2018.
46. Huikuri HV, Castellanos A, Myerburg RJ. Sudden death due to cardiac arrhythmias. New England Journal of Medicine. 2001 Nov 15;345(20):1473-82.
47. Kadambe S, Murray R, Boudreaux-Bartels GF. Wavelet transform-based QRS complex detector. IEEE Transactions on biomedical Engineering. 1999 Jul;46(7):838-48.
48. Ye C, Kumar BV, Coimbra MT. Heartbeat classification using morphological and dynamic features of electrocardiogram signals. IEEE Transactions on Biomedical Engineering. 2012 Aug 15;59(10):2930-41.
49. Addison PS. Wavelet transforms and the electrocardiogram: a review. Physiological measurement. 2005 Aug 8;26(5): R155.
50. Sahoo P, Riedel T. Mean value theorems and functional equations. World Scientific; 1998.
51. Retrieved from http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf. Accessed on 11/08/2017
52. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine. 2001 May;20(3):45-50.